

Supplemental material

April 12, 2021

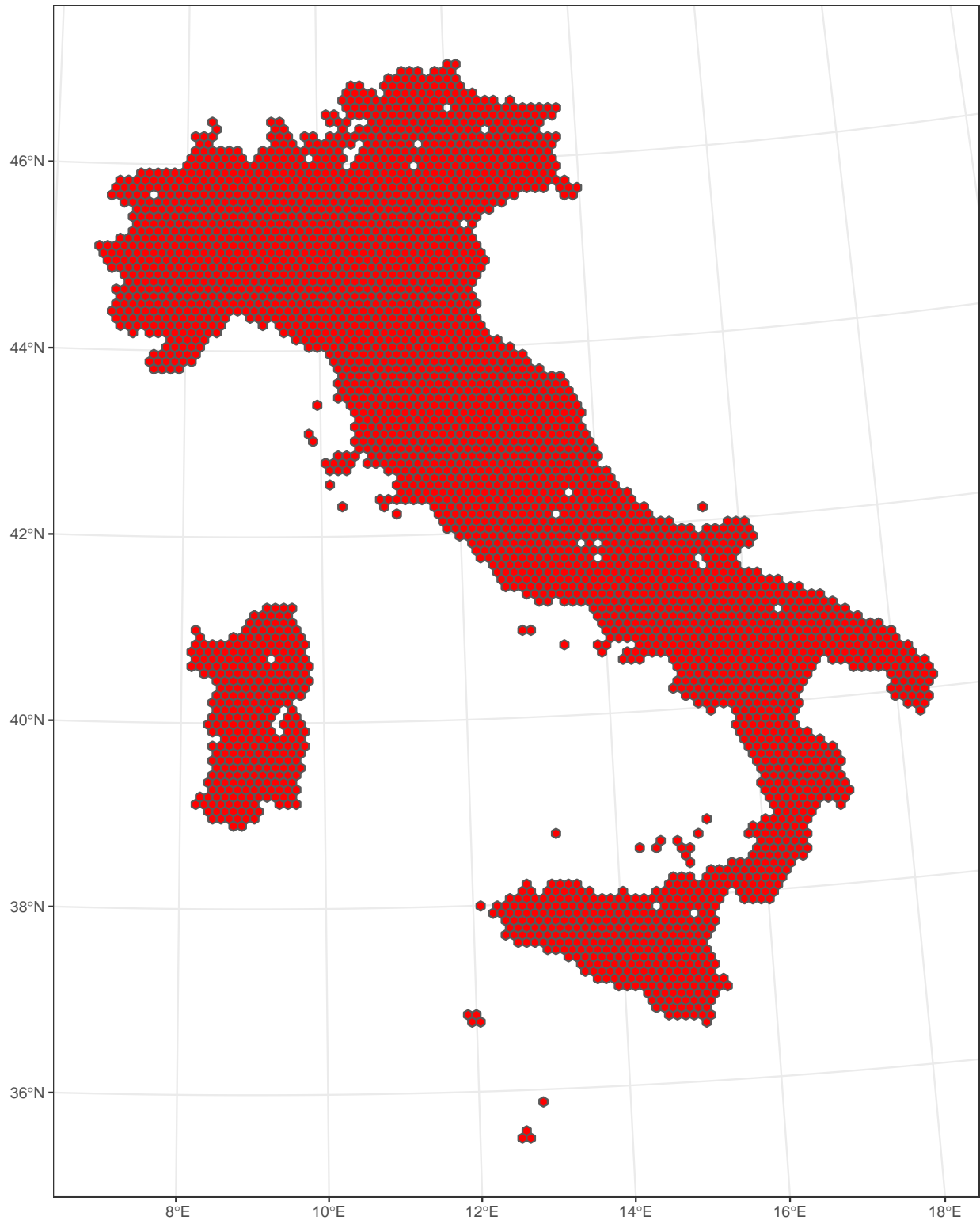
Contents

1	Hexagonal Grid	2
1.1	Interesection between census sections and hexagonal cells for count variables	4
1.2	Interesection between comune and hexagonal cells for count variables	4
1.3	Intersection between areas and hexagon cells for continuous variables	5
1.4	Intersection between variables reported at the regional level and hexagon cells	5
2	Variables	5
2.1	Excess mortality	5
2.2	Air quality variables	9
2.3	Atmosphere variables	14
2.4	Census variables	14
2.5	Testing variables	15
2.6	Lockdown	15
2.7	Commuter-weighted network	17
2.8	Neighboring excess mortality	20
2.9	Road density	20
2.10	Hotel rooms	21
2.11	Icu beds	21
3	Modelling	22
3.1	Standardised coefficient table	22
3.2	Standardised coefficient plot	24
3.3	Predictions for the Province of Bergamo	25

1 Hexagonal Grid

An hexagonal grid is randomly created to cover the entire Italian territory. Each grid is set to have a size (diameter) of 10-km. Variables associated to each cell are used as the unit of analysis for the regression models. The adoption of a grid of hexagonal cell is justified by the necessity of mapping statistics reported for geographies with significantly different shape and area.

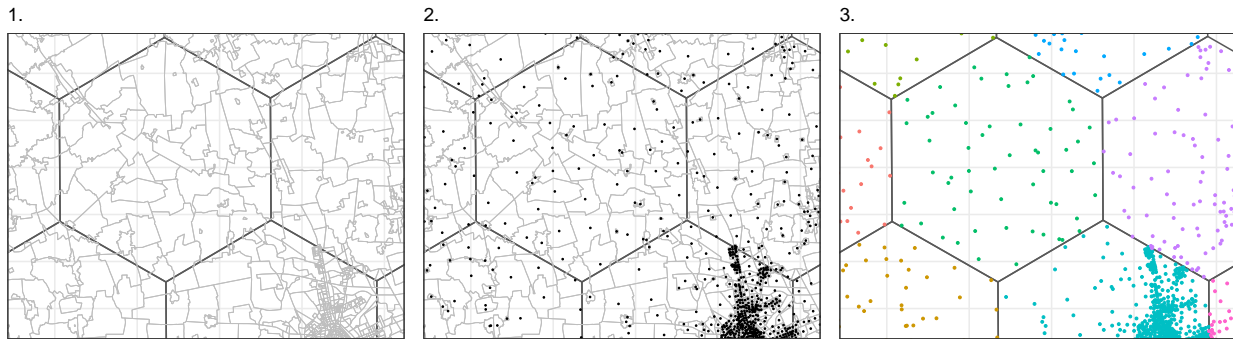
The following figure shows the grid of 3801 hexagonal cell used in the analysis. Notably, 271 cells have been dropped since based on the 2011 census are not inhabited.



1.1 Interesection between census sections and hexagonal cells for count variables

To map census count variables measured at the level of census sections with hexagonal cells, we proceeded as following:

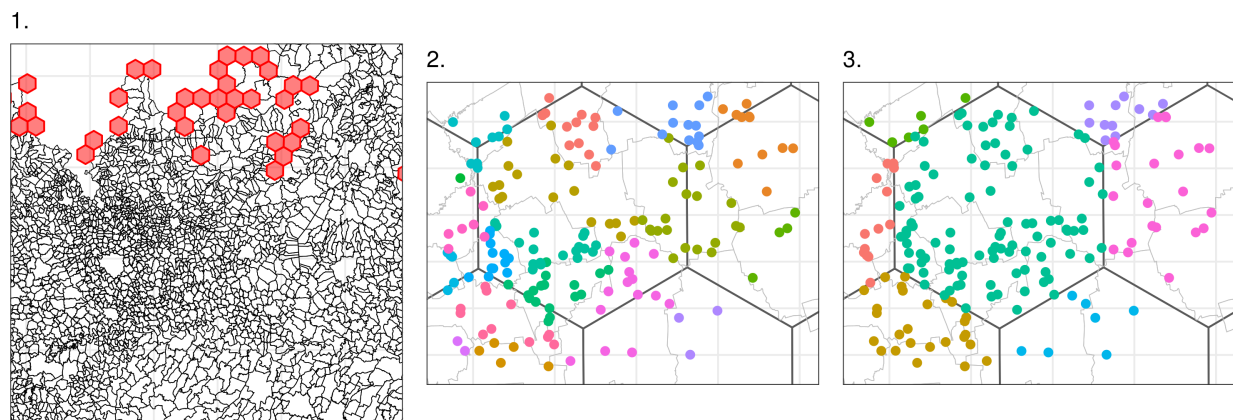
1. we computed the centroid, or its geographic centre, of the 402,678 census sections;
2. we intersected centroids and hexagonal cells;
3. for each hexagonal cell, we summed the count variables pertaining to the census sections with a centroid contained within the cell.



1.2 Interesection between comune and hexagonal cells for count variables

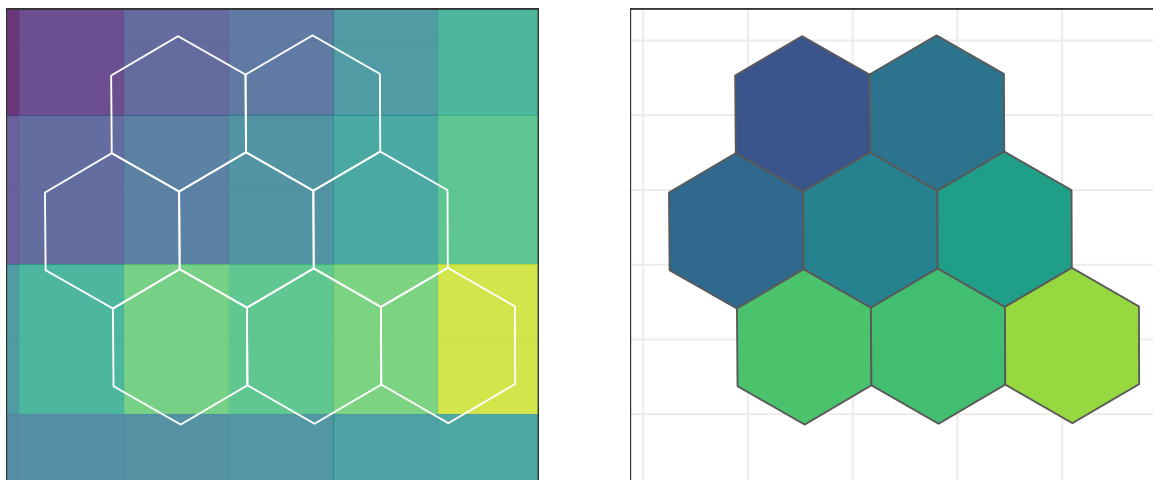
To map *comune* count variables onto hexagon cells, we proceeded as following:

1. we clipped unpopulated areas corresponding to the unpopulated hexagon cells identified before from the *comune*;
2. we randomly sampled a number of points within the remaining (populated) area of the *comune* corresponding to the reported count;
3. we counted how many points resulted within each hexagonal cell.



1.3 Intersection between areas and hexagon cells for continuous variables

When continuous variables were reported for a grid of cells, as with air quality and atmospheric data, or when count variables were reported at the level of comune for multiple income brackets, as in the case of income data, the variable for each hexagonal cell was computed as the average of the values reported for the intersecting cells (or comune) weighted according the extension of the area of intersection.



1.4 Intersection between variables reported at the regional level and hexagon cells

For variables reported at the regional level, each hexagon cell assumes the value reported for the region where it is located. In the case of a hexagonal cell intersecting more than one region, the cell is assigned to the region with the largest intersection area.

2 Variables

2.1 Excess mortality

Excess mortality is based on Istat figures published in March 2021 ([link](#)).

Figures for deaths recorded in every calendar day for the period 2015-19 are reported for 7903 *comune*. Figures for deaths recorded for each calendar day for the first nine months of 2020 are reported for 7903 *comune* (or 100% of the total). In 2019, these 7903 *comune* reported 100% of the number of deaths reported in Italy.

A video with the excess mortality mapped in Italy for 2020 is available here <https://youtu.be/ZuHpZbsbqIY>.

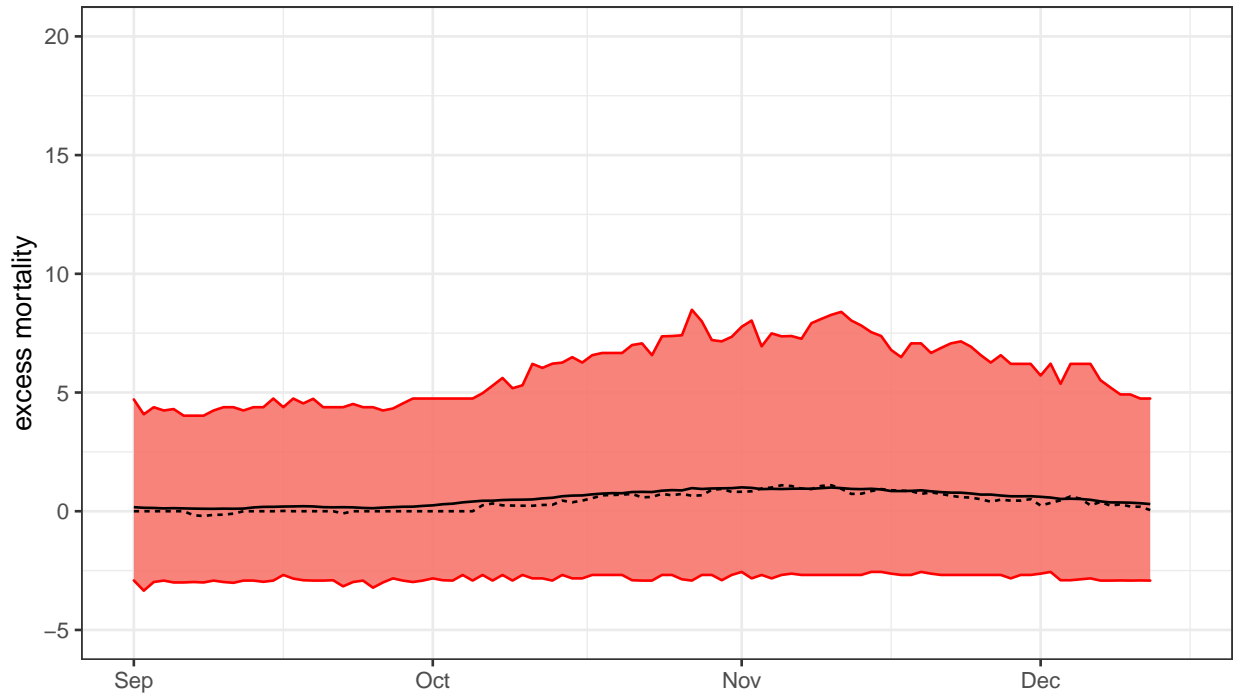
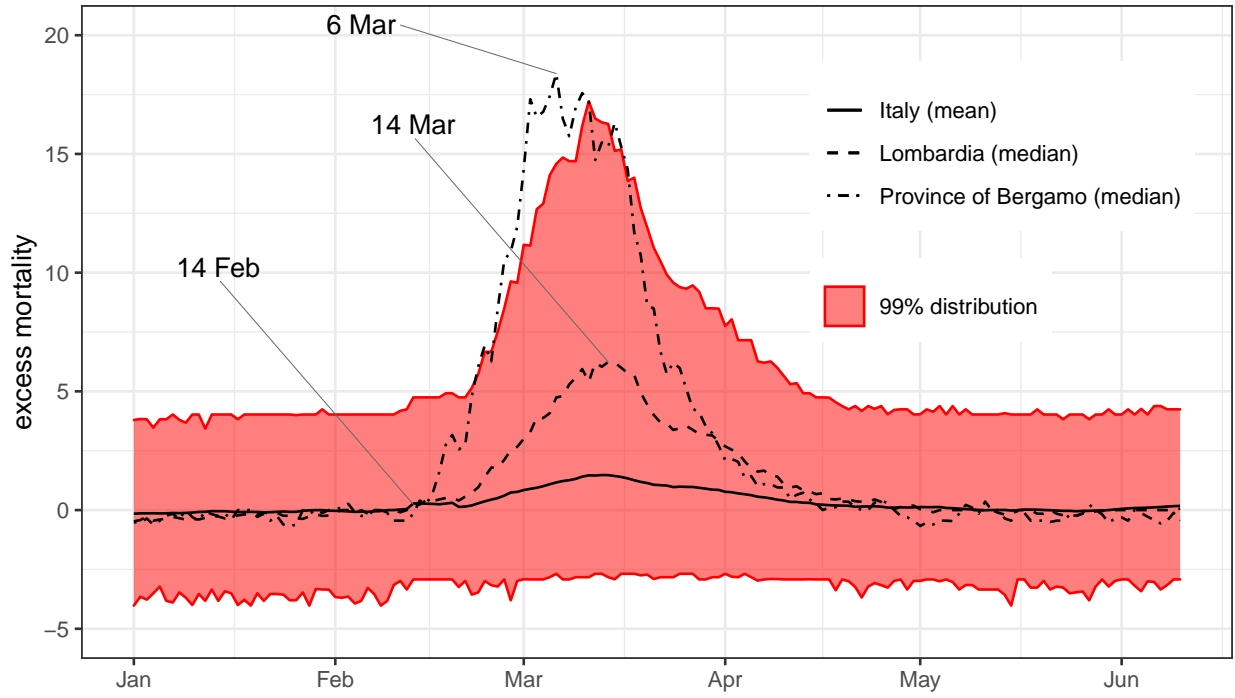
We compute excess mortality through the following steps:

1. The daily number of deaths reported for each *comune* between 2015 and 2020 is mapped onto our hexagon cells with the methods detailed above for counted variables.
2. For each year between 2015 and 2020, we calculated for each hexagon cell a centered 7-day moving average of the number of reported deaths over 65 in each comune for each year from 2015 to 2020.
3. For each calendar day, the mean and the standard deviation of the moving averages from 2015 to 2019 are calculated.

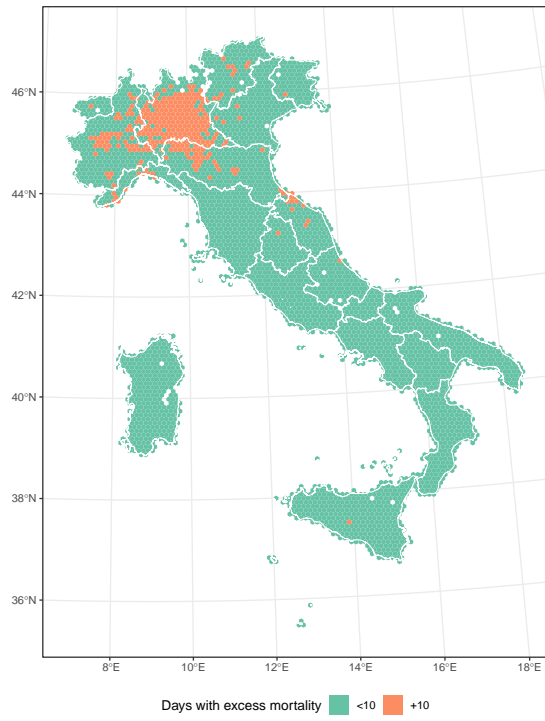
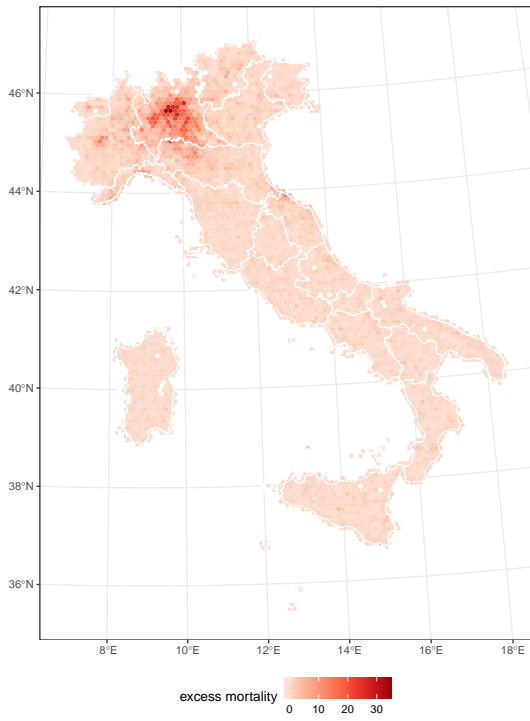
4. For each calendar day, the difference between the 2020 moving average and the 2015-2019 mean is divided by the 2015-2019 standard deviation. In doing so, we define the standard deviation as the unit of measurement and normalise differences across cells due to population density.
5. We shifted the resulting time series backward between 14 and 17 days to the estimated day of the infection resulting in death instead of the actual day of reported death. This takes into consideration changes in the median number of days between symptoms' onset and death in Italy. Based on data published by Italy's Italian National Health Service (Istituto Superiore di Sanità) we set the contact-to-death estimate to
 - 14 days before 13 April 2020,
 - 15 days before 28 May 2020,
 - 16 days before 9 July 2020,
 - 17 days on and after 9 July 2020.
6. Infinite and non-number values resulting from division by zero are assigned an excess mortality value of zero.

With this method we are able to estimate the excess mortality for 4072 hexagon cells and for each day between 01 January 2020 and 12 December 2020.

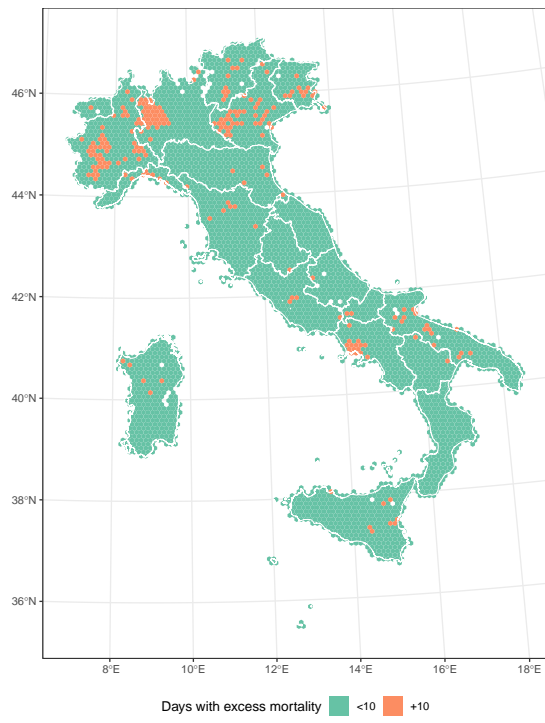
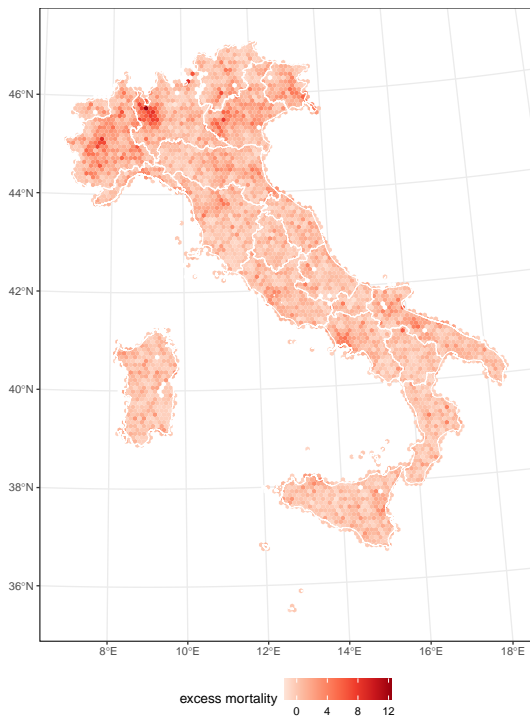
The distribution of the resulting excess mortality variable is detailed over time and over space in the following figures.

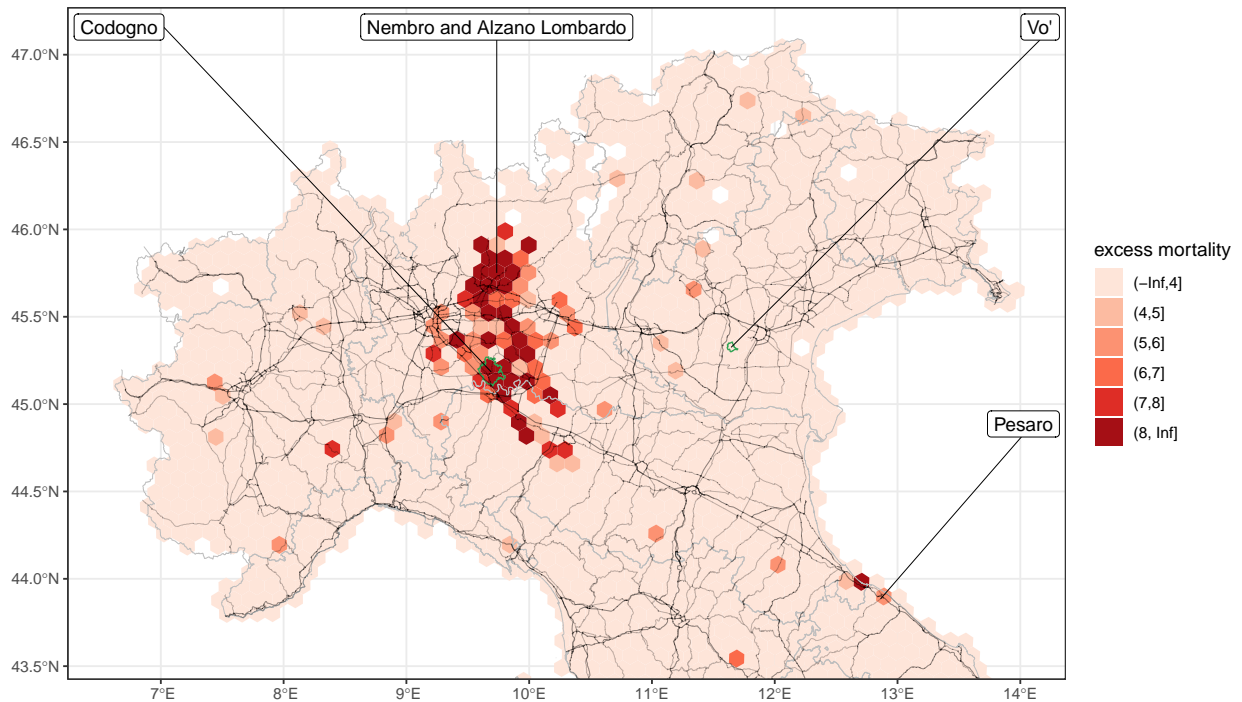
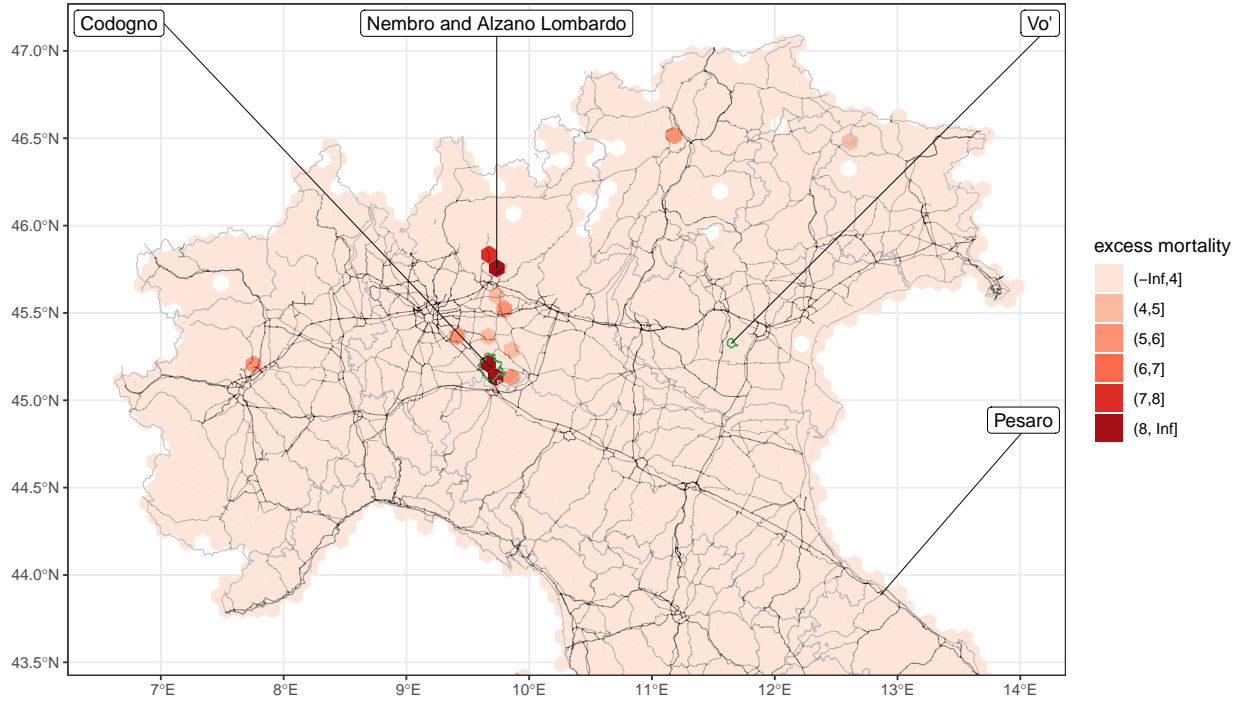


First wave (Feb – Apr 2020)



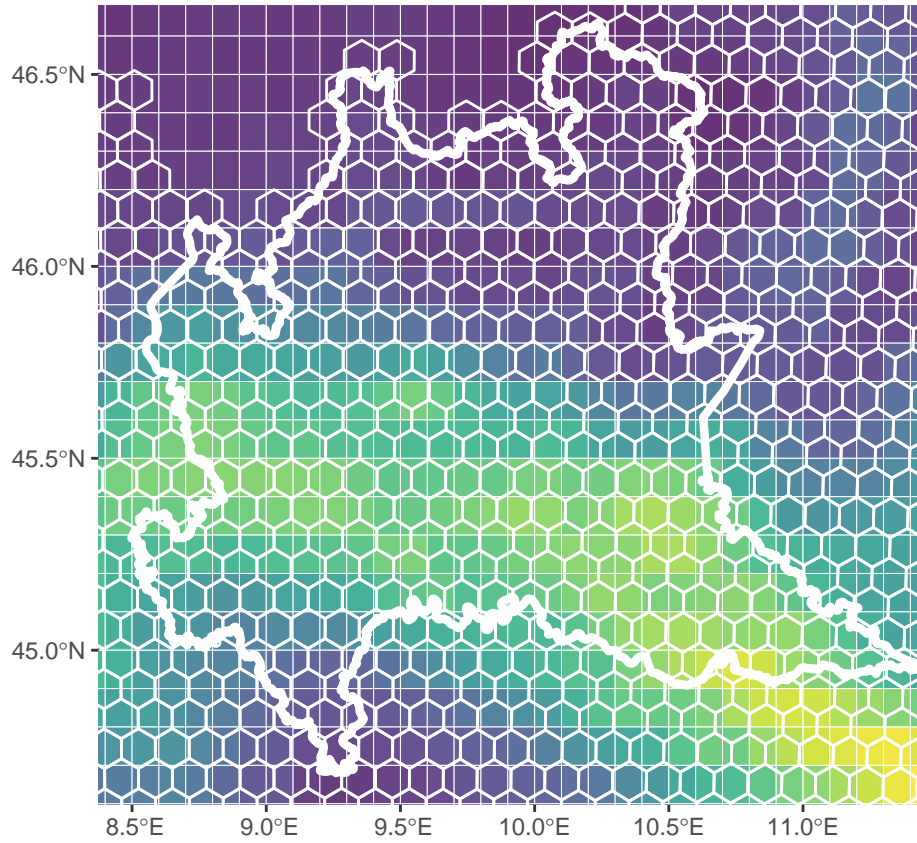
Second wave (Oct – Dec 2020)



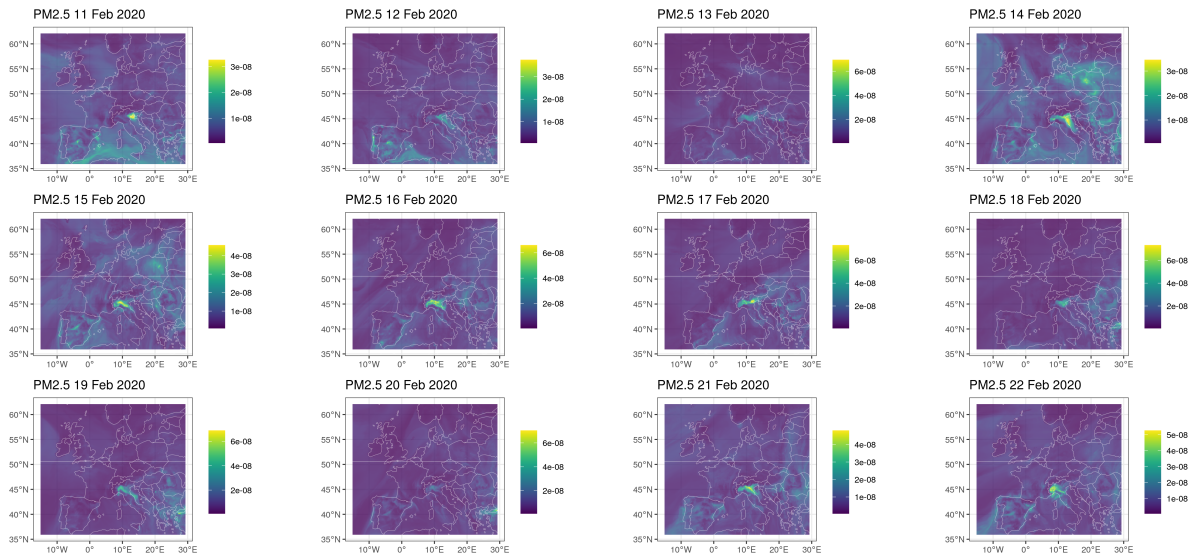


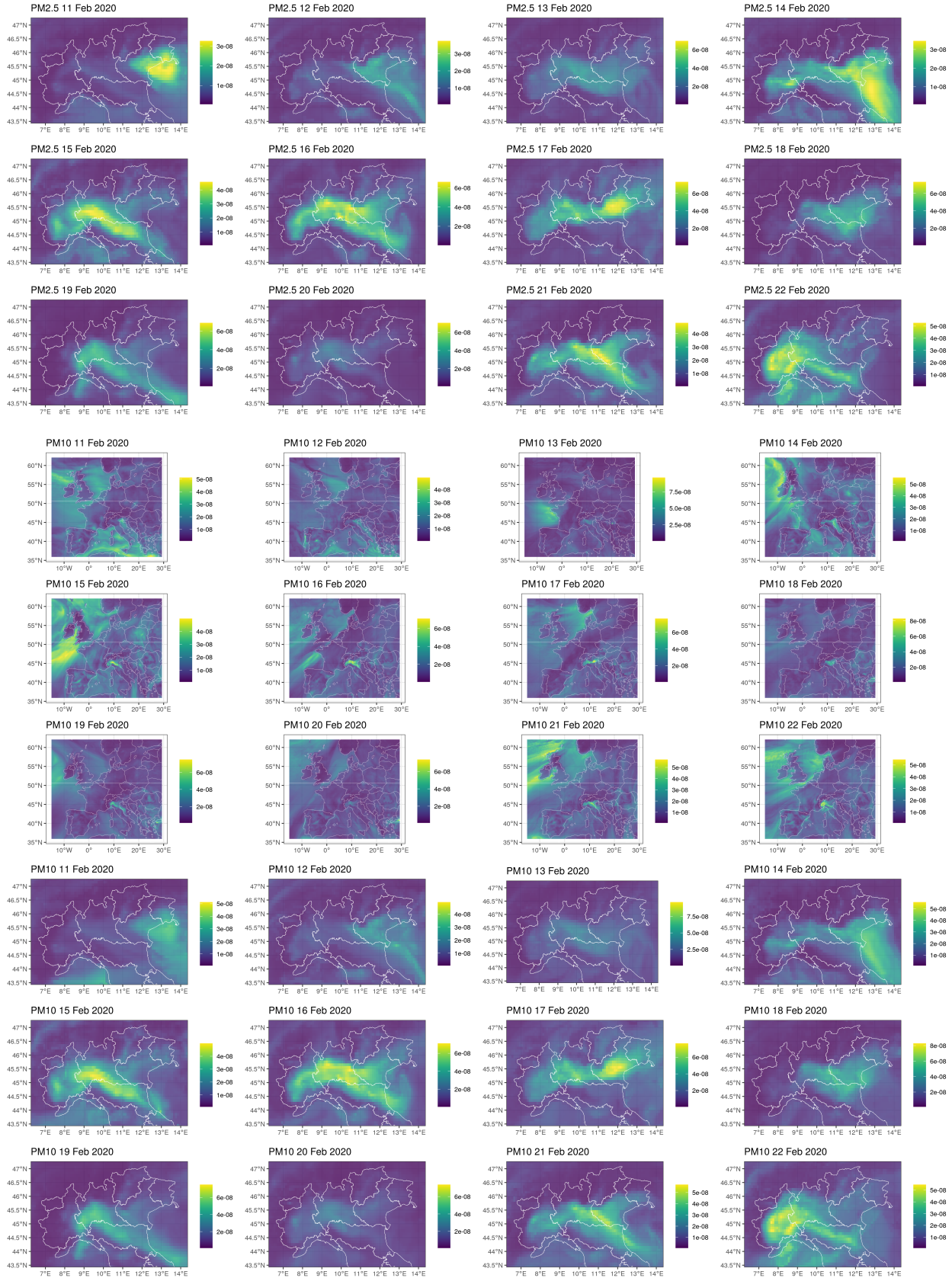
2.2 Air quality variables

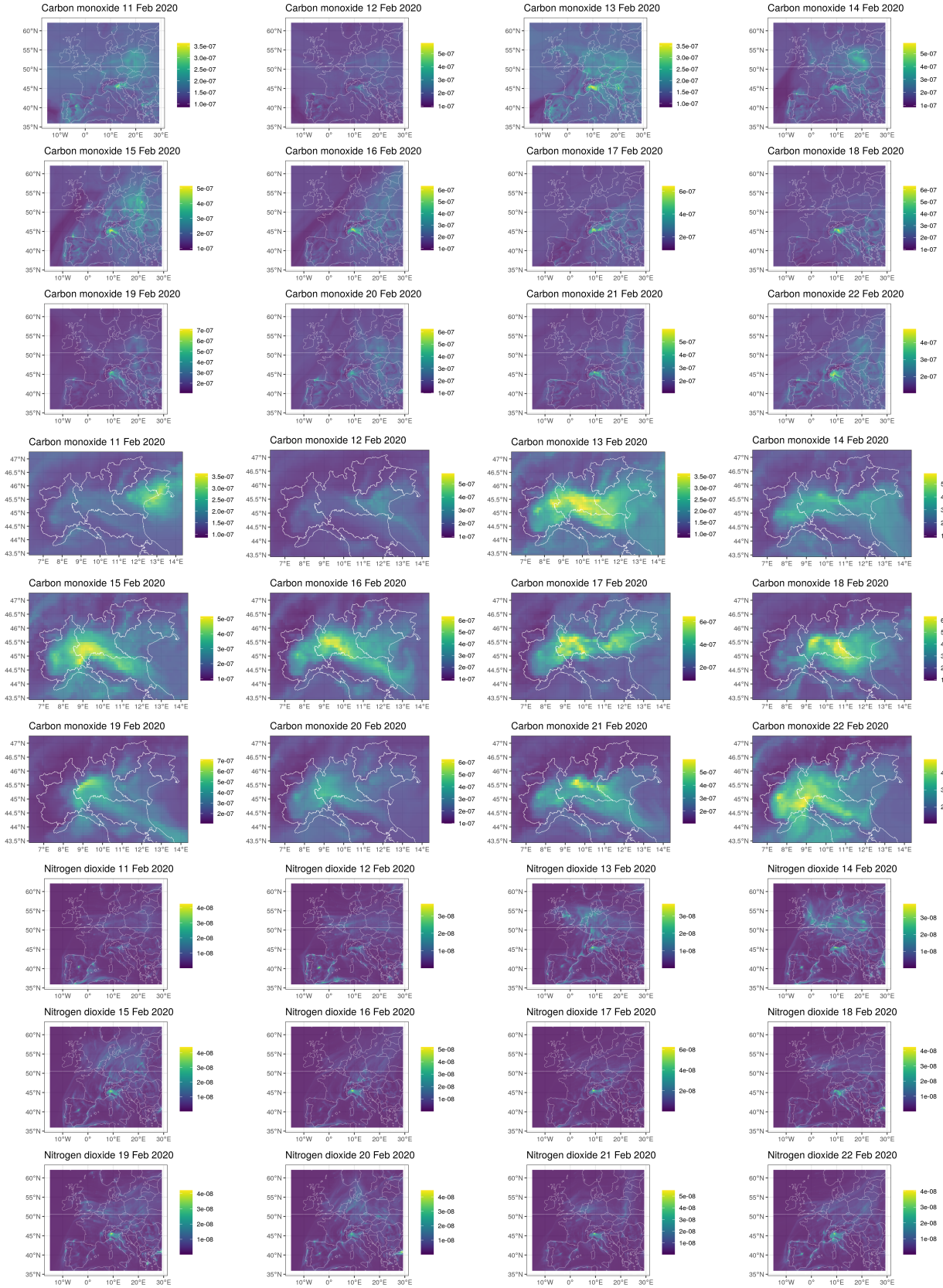
Air quality data are obtained from Copernicus Atmosphere Monitoring Service (CAMS, [link](#)) with a 10 kilometres' resolution. In the following figures, CAMS cells are overlaid to our hexagon cells.

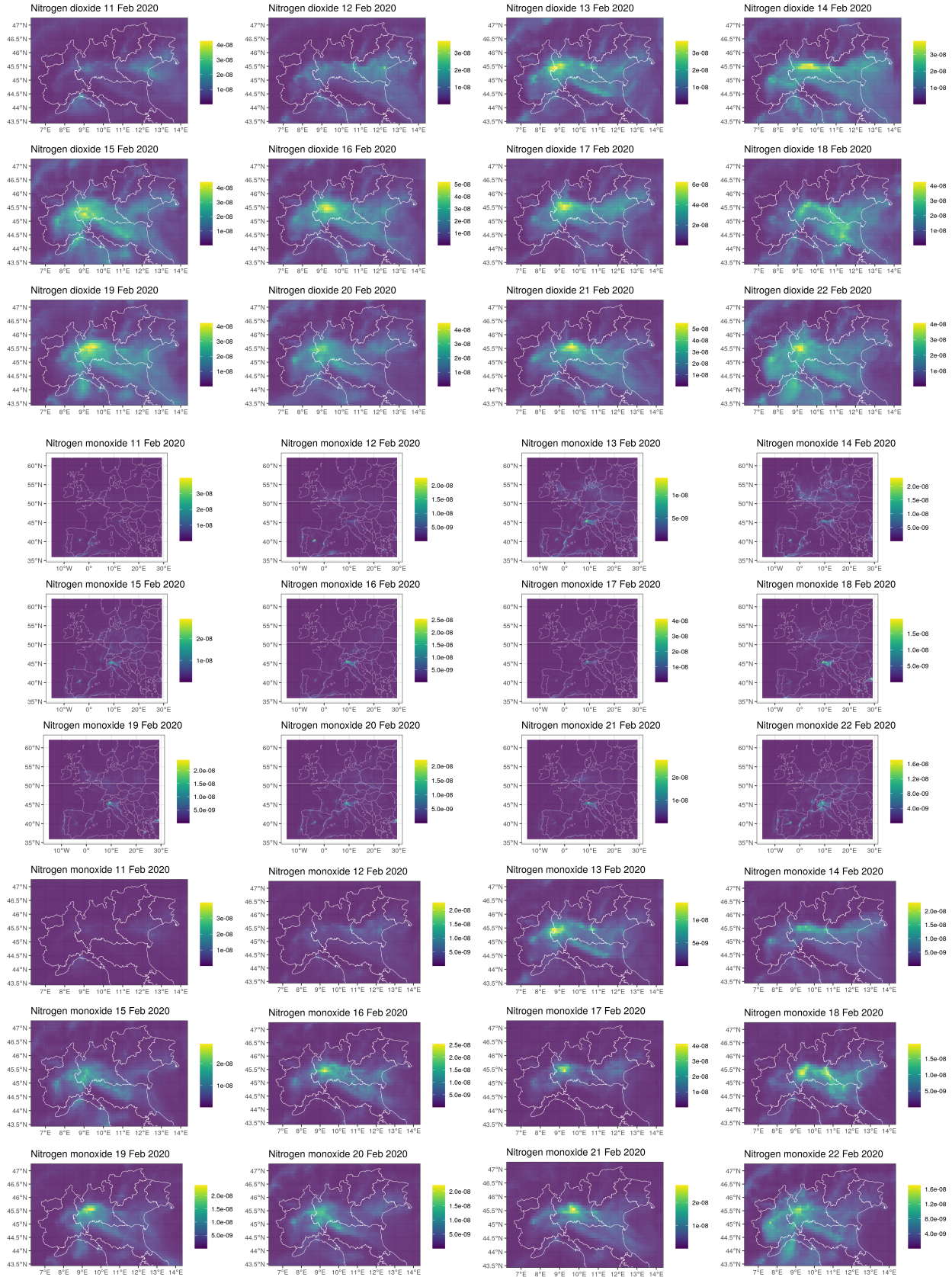


In the following figures we map instead the evolution of selected air quality measures over Europe and Northern Italy.



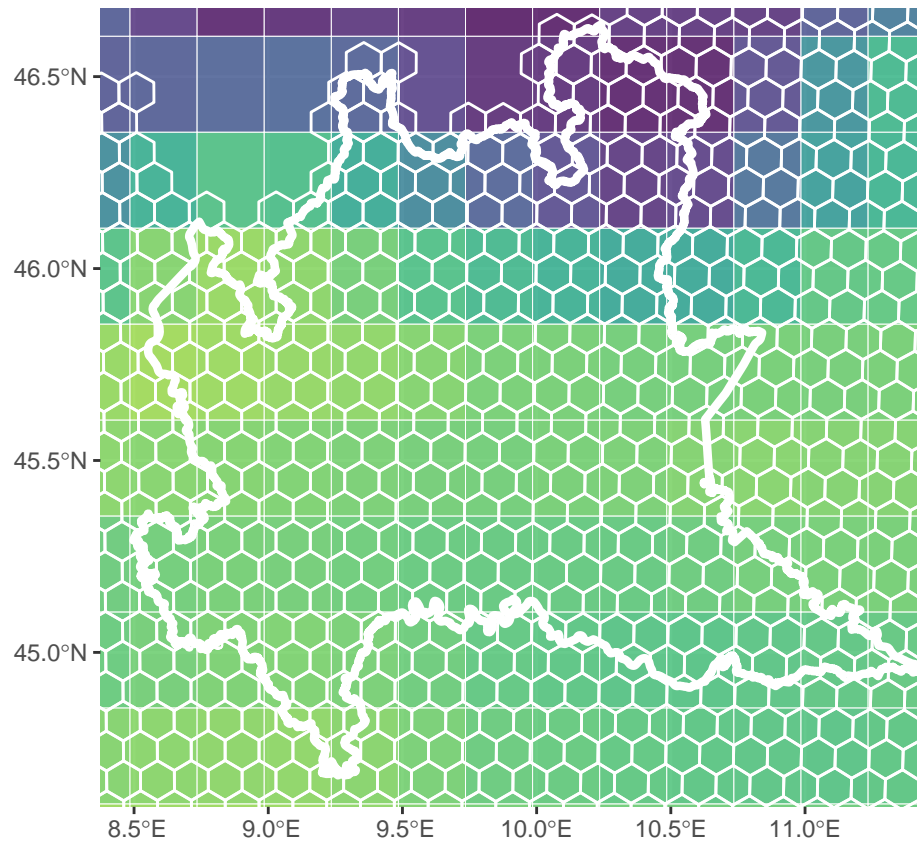






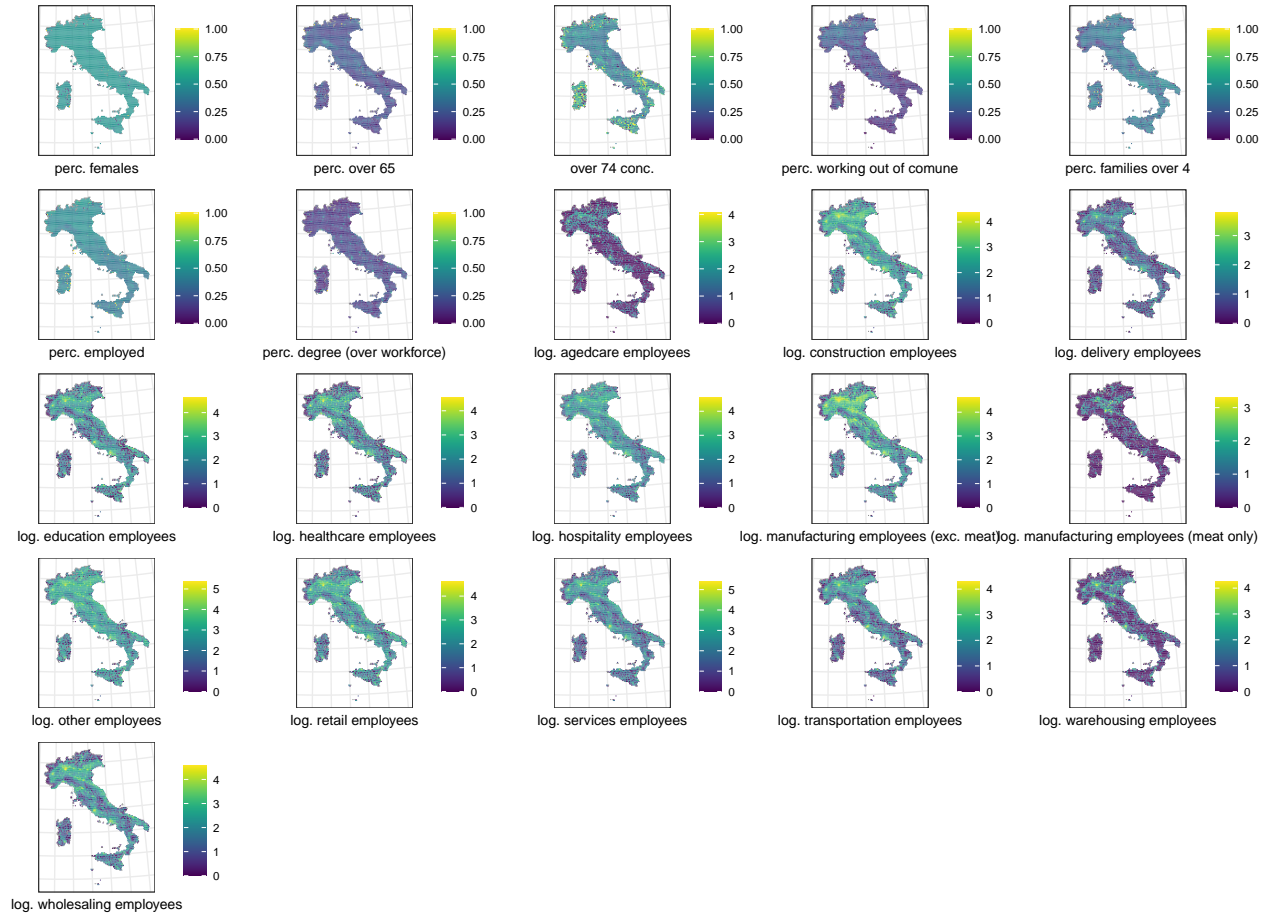
2.3 Atmosphere variables

Atmosphere data are derived from ERA5 (link) variables with a 30 kilometres' resolution. In the following figures, ERA5 cells are overlaid to our hexagon cells.



2.4 Census variables

Demographic variables are computed based on census counts at the level of 2011 census sections. The geographic distribution of the variables used in the model is showed in the following figures.



2.5 Testing variables

Testing data are based on official daily figures published by the Italian government ([link](#)) and reported at the level of region. The first day for which figures are available is 24 February 2020. The case-to-test ratio is computed dividing the 7-day center-aligned moving average of number of positive tests divided by the 7-day center-aligned moving average of the number of tests.

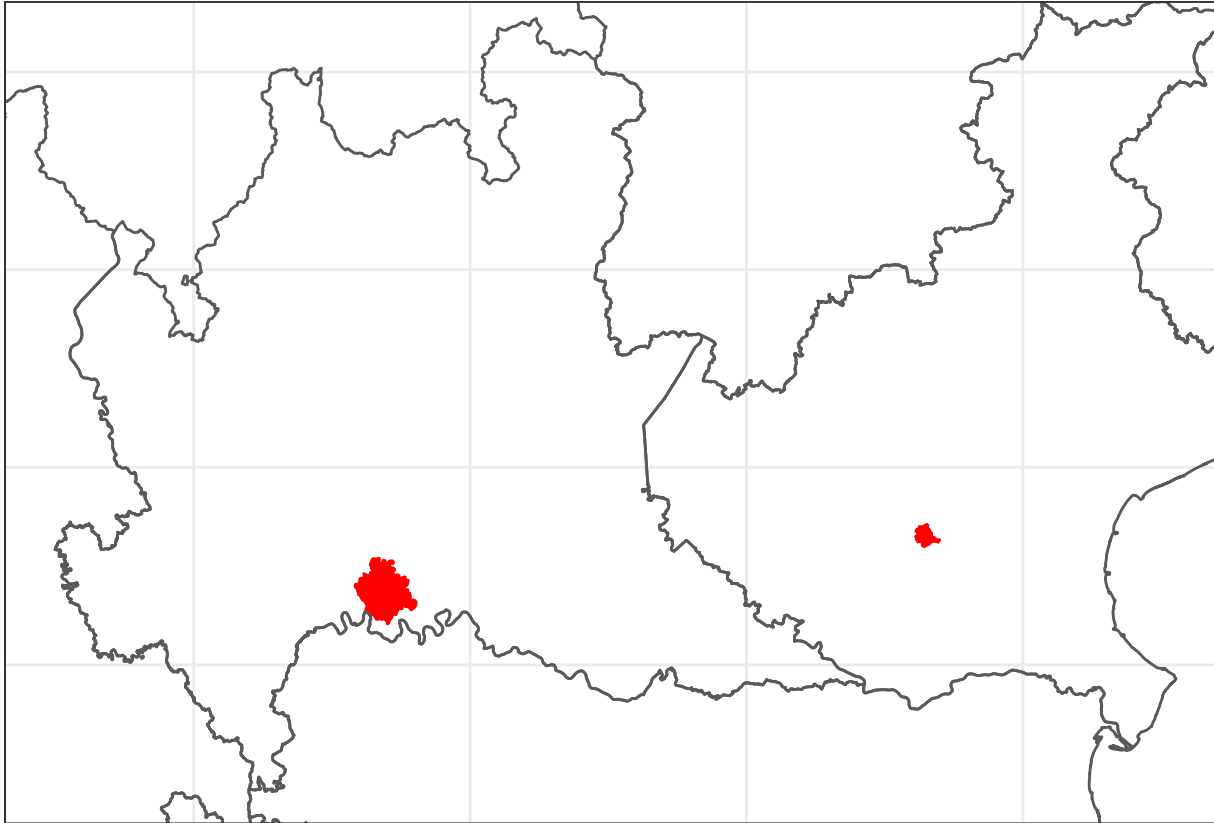
The following figure shows the evolution of the case-test ratio in the 20 Italian regions.

2.6 Lockdown

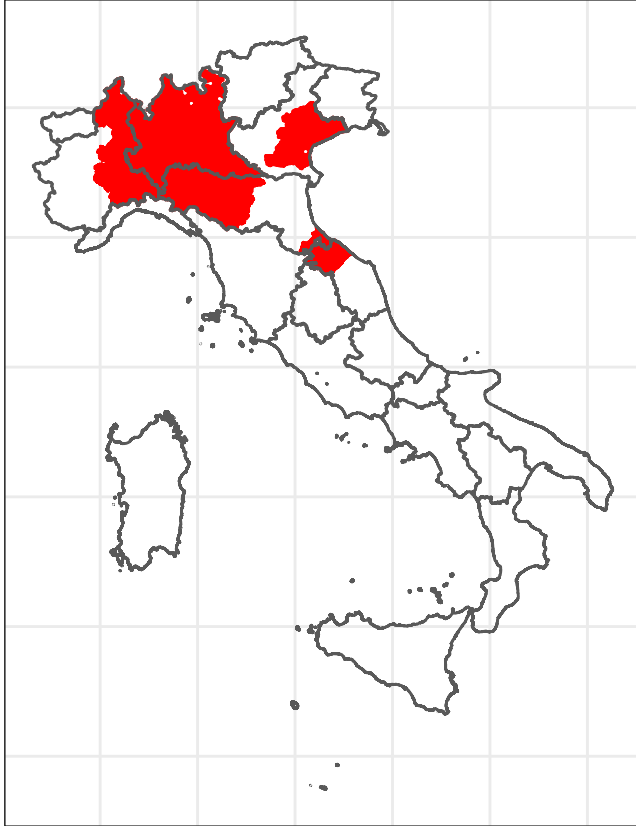
For each hexagon cell, the percentage of the population estimated to be living under lockdown measures is computed through the following steps:

1. The areas resulting from the union of the area of the *comune* targetted by the lockdown measures established between 23 February and 31 December 2020.
2. For each hexagon cell the percentage of the population under lockdown measures was estimated to be 0% (100%) if the area of the cell was included (excluded) entirely by the two lockdown areas.
3. When an hexagon cell was only partially covered by one of the two lockdown areas, the percentage of the population under lockdown measures was estimated by
4. summing the count of the population of all the census sections pertaining to a *comune* under lockdown measures and, based on the census section's centroid, intersecting with the hexagon cell.
5. dividing the resulting sum for the entire population estimated to live in the area of the hexagon cell (see above).

The following *comune* were subjected to lockdown measures on 23 February 2020: .



All the *comune* in Lombardy and in the the following *provincia*: Alessandria, Asti, Modena, Novara, Padua, Parma, Piacenza, Pesaro and Urbino, Reggio Emilia, Rimini, Treviso, Venice, Verbano-Cusio-Ossola, were subjected to lockdown measures on 08 March 2020.



2.7 Commuter-weighted network

A network connecting the hexagon cells and weighted based on the number of people commuting between pairs of cells is estimated based on a *comune-to-comune* asymmetric matrix populated with the count of commuters on census day 2011.

Consistently with the approach followed in computing other variables, we proceeded as following:

1. For each *comune*, we computed how its 2011 population distributes across different hexagonal cells so to have the following table

Comune ID	Hexagonal cell ID	Population prop.
1	1	50%
1	2	50%
2	1	70%
2	2	30%
3	3	10%
3	4	10%
3	5	80%

2. We joined the resulting table with the long version version of the asymmetric matrix, so to have the following table

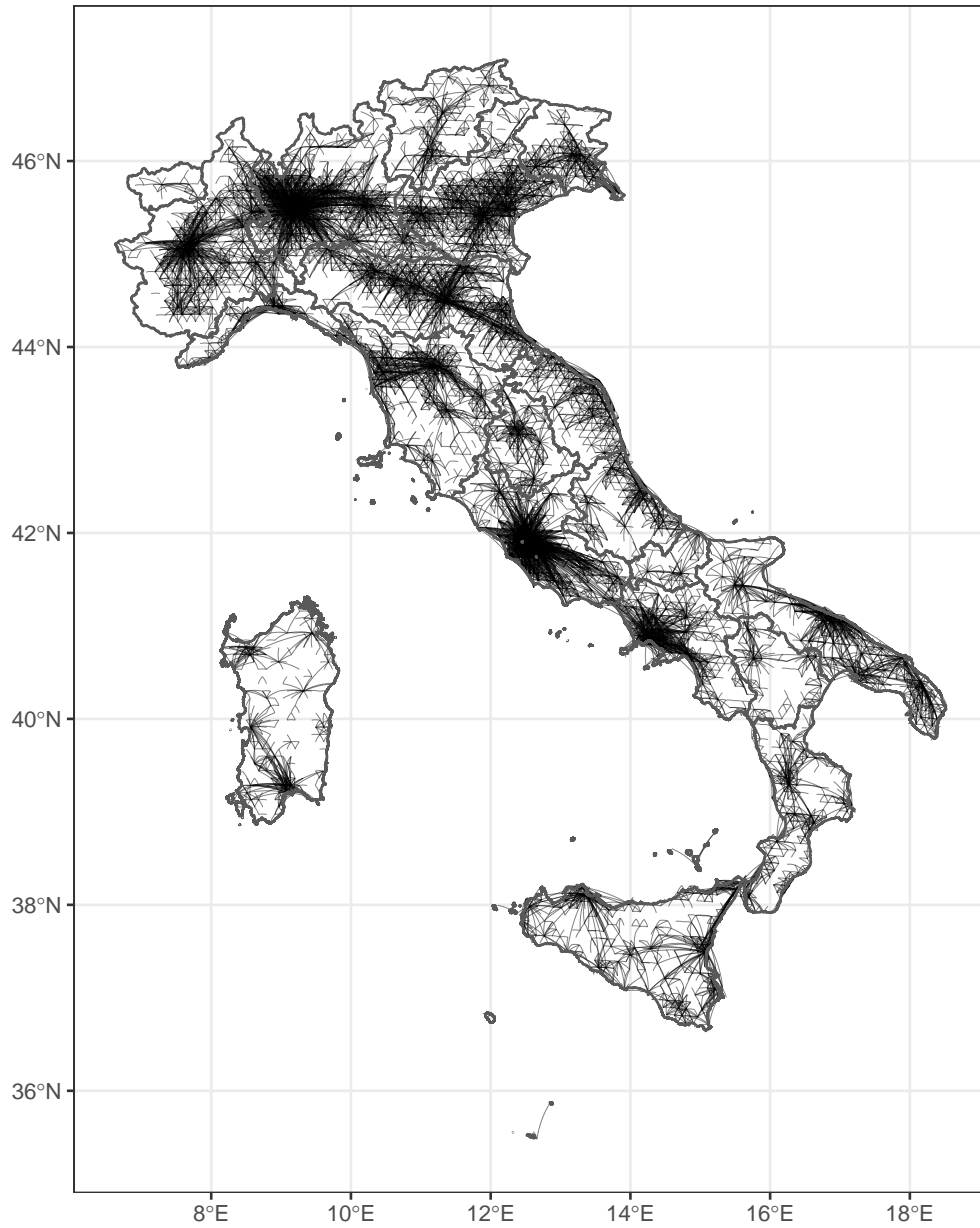
From comune ID	From hexagonal cell ID	From population prop.	To comune ID	N. commuters
1	1	50%	2	10
1	1	50%	2	10

From comune ID	From hexagonal cell ID	From population prop.	To comune ID	N. commuters
1	1	50%	2	10
1	1	50%	2	10
2	1	70%	3	15

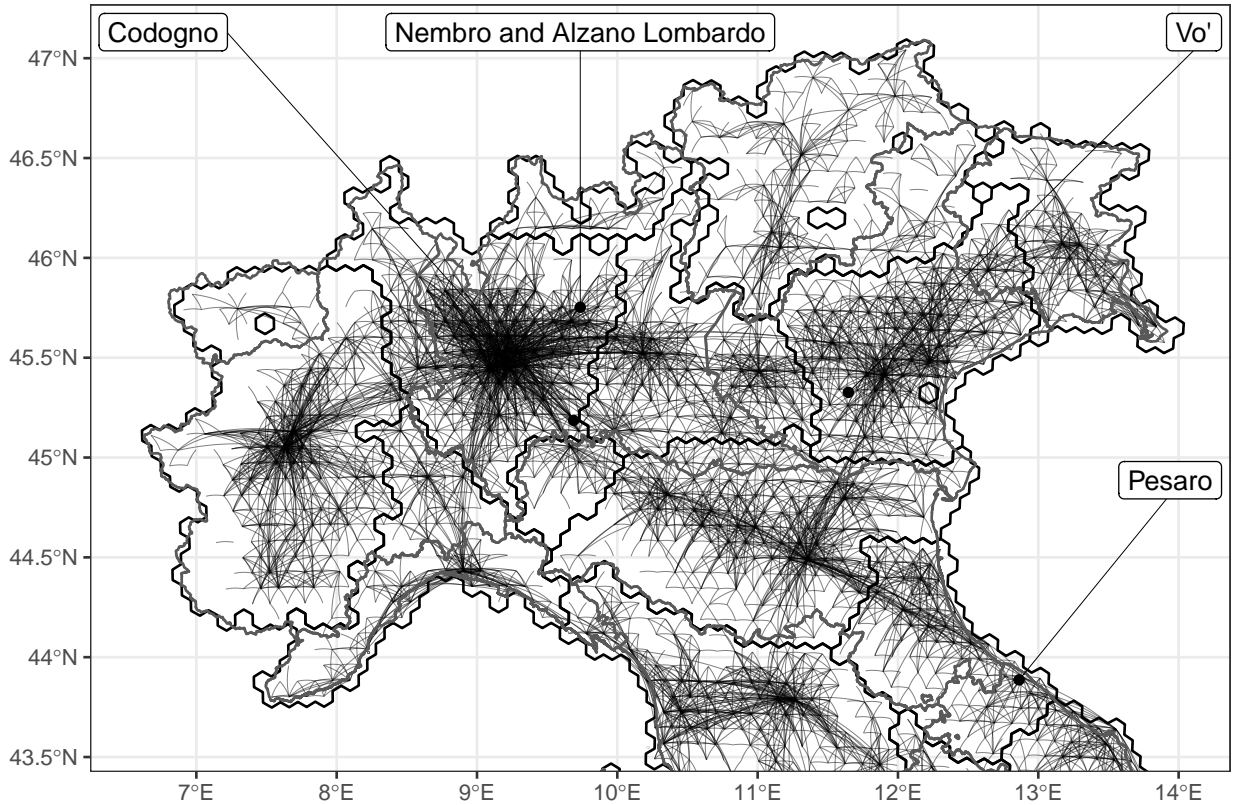
3. We combined this last table with the table generated in the first step joining on `to Comune ID = Comune ID` resulting in

...	From hex. ID	From pop. %	...	N. commuters	To hex. ID	To pop. %
1	1	50%	2	10	2	30%
1	1	50%	2	10	3	10%
1	1	50%	2	10	4	20%
1	1	50%	2	10	5	10%
2	1	70%	3	15	4	100%

4. For each row of the resulting table, we then computed `weighted commuting = n. commuters × from pop. % × to pop. %` as the relationship between a pair of hexagonal cells defined by the commuting from $comune_A$ to $comune_B$.
5. For each pair of hexagon cells A and B , we sum all the resulting `weighted commuting` to obtain `weighted commuting $_{\vec{AB}}$` (which is different from `weighted commuting $_{\vec{BA}}$`).
6. We created direct network mapping the asymmetric relationships between pairs of hexagonal cells, where each relation was assigned weight `weighted commuting $_{\vec{AB}}$`
7. We simplify the directed network to obtain the undirected network with symmetrical relationship between pairs of hexagonal cell and weight `weighted commuting $_{\overline{AB}}$` resulting from `weighted commuting $_{\vec{AB}}$ + weighted commuting $_{\vec{BA}}$` illustrated by the following Figure.



The following Figure shows the same network and the communities identified by a community detection algorithm (A Clauset, MEJ Newman, C Moore: Finding community structure in very large networks, <http://www.arxiv.org/abs/cond-mat/0408187>).



2.8 Neighboring excess mortality

We computed neighboring excess mortality based on the commuter-weighted network detailed above. For each day and for each cell, neighboring excess mortality was the average excess mortality in the other cells on that day weighted according to the corresponding $\text{weighted commuting}_{AB}$.

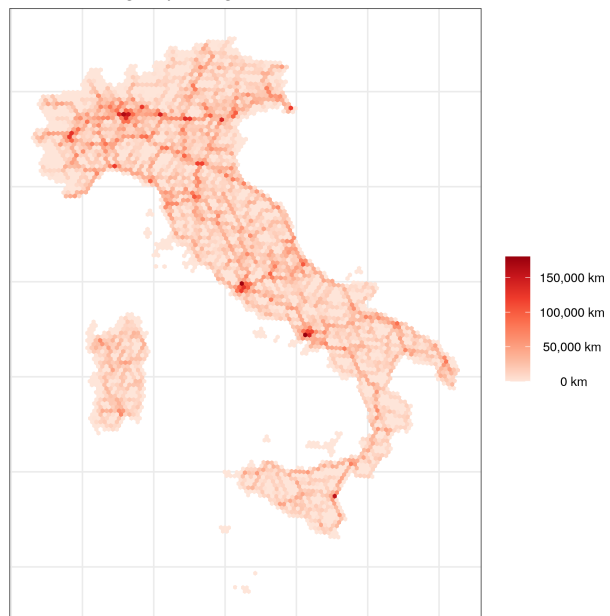
2.9 Road density

The road density variable is computed based on the length of primary roads and based on OpenStreetMap data. For each hexagon cell, the total length of is computed by summing all the road segments contained in the cell. Original data and length by cell is reported in the following figure.

OpenStreetMap primary road network



Total road length by hexagon cell

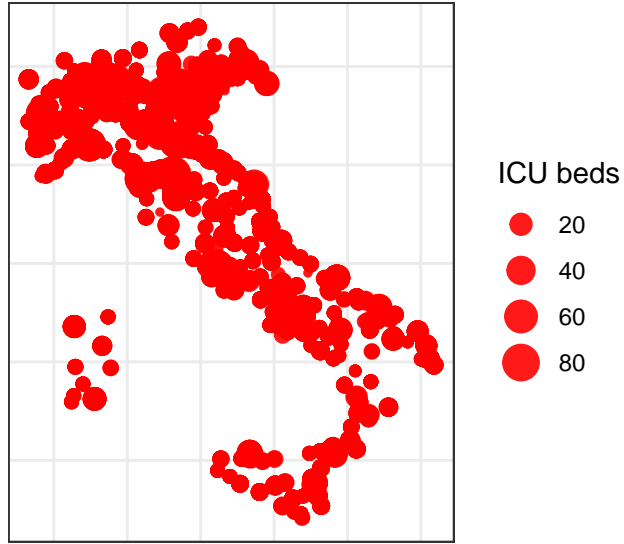


2.10 Hotel rooms

For each cell, we computed the number of hotel rooms based on data reported by Istat at the level of *comune* and according to the mapping method for count variables illustrated above.

2.11 Icu beds

The number of ICU beds (“terapia intensiva”) was computed based on a list of health centres compiled by the Ministry of Health and containing the street address. From that list, we geocoded 2147 street address with at least one ICU beds using the Google Maps API. The distribution of ICU beds is mapped in the following Figure.



3 Modelling

3.1 Standardised coefficient table

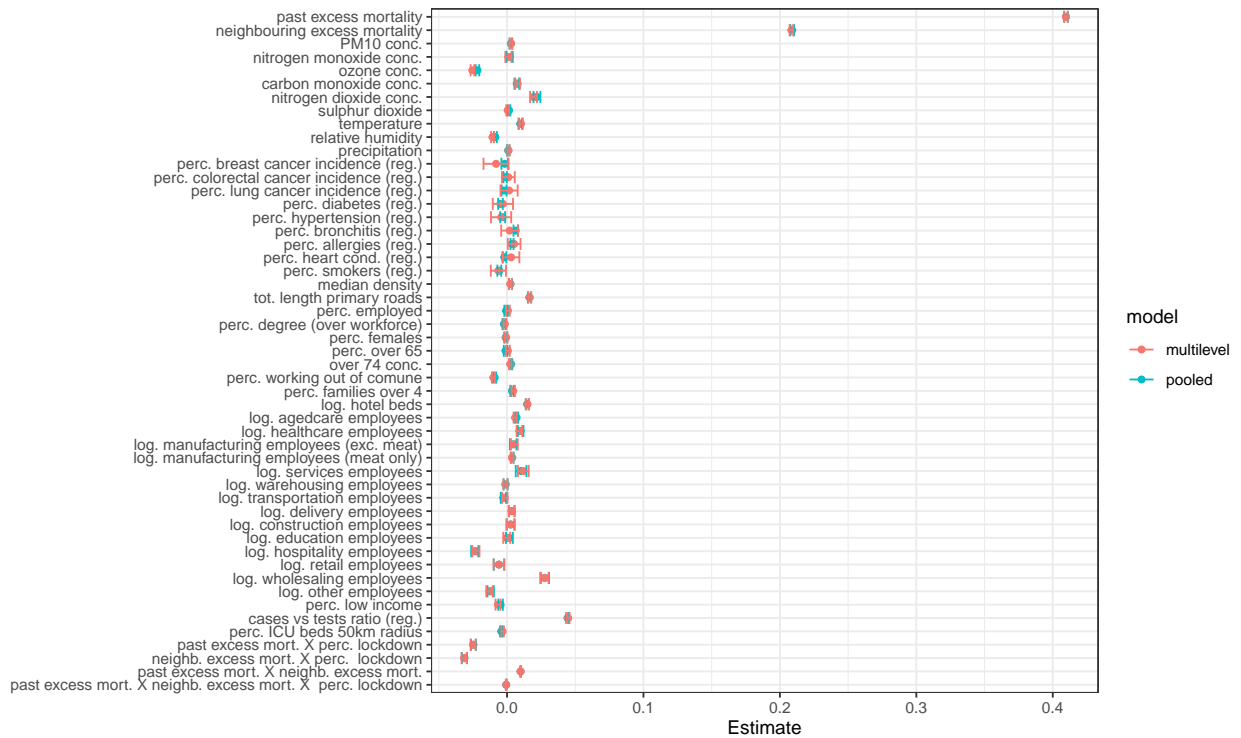
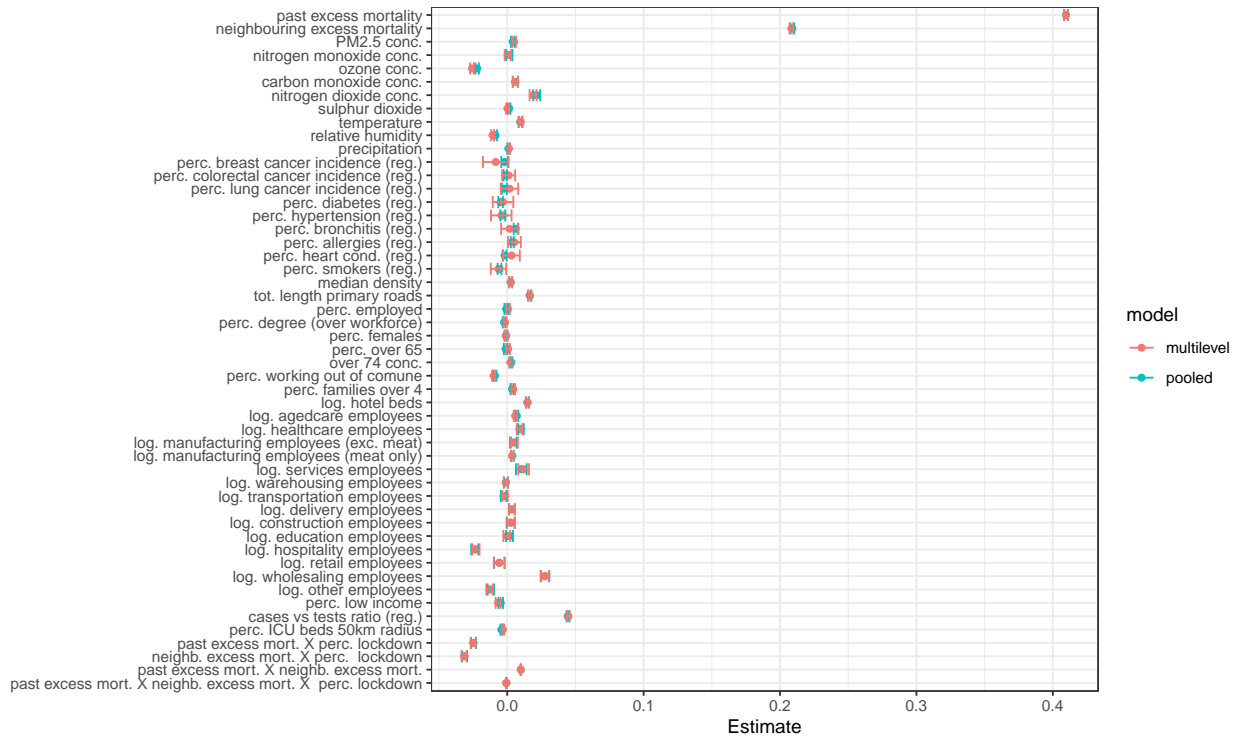
Table 4:

	<i>Dependent variable:</i>			
	Excess mortality (backdated by 17 days)			
	<i>OLS</i>		<i>linear mixed-effects</i>	
	(1)	(2)	(3)	(4)
past excess mortality	0.410*** (0.001)	0.410*** (0.001)	0.410*** (0.001)	0.410*** (0.001)
perc. lockdown	-0.017*** (0.002)	-0.016*** (0.002)	-0.016*** (0.002)	-0.016*** (0.002)
neighbouring excess mortality	0.209*** (0.001)	0.209*** (0.001)	0.208*** (0.001)	0.208*** (0.001)
PM2.5 conc.	0.004*** (0.001)		0.005*** (0.001)	
PM10 conc.		0.003*** (0.001)		0.003*** (0.001)
nitrogen monoxide conc.	0.001 (0.002)	0.002 (0.002)	0.0004 (0.002)	0.001 (0.002)
ozone conc.	-0.022*** (0.001)	-0.022*** (0.001)	-0.026*** (0.001)	-0.025*** (0.001)
carbon monoxide conc.	0.006*** (0.002)	0.007*** (0.002)	0.006*** (0.002)	0.008*** (0.002)
nitrogen dioxide conc.	0.022*** (0.003)	0.022*** (0.003)	0.019*** (0.003)	0.019*** (0.003)
sulphur dioxide	0.001 (0.001)	0.001 (0.001)	0.0002 (0.001)	0.0005 (0.001)
temperature	0.010*** (0.001)	0.010*** (0.001)	0.010*** (0.001)	0.010*** (0.001)
relative humidity	-0.009*** (0.001)	-0.009*** (0.001)	-0.011*** (0.001)	-0.011*** (0.001)
precipitation	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
perc. breast cancer incidence (reg.)	-0.002 (0.002)	-0.002 (0.002)	-0.008 (0.009)	-0.008 (0.009)
perc. colorectal cancer incidence (reg.)	-0.001 (0.001)	-0.001 (0.001)	0.001 (0.005)	0.001 (0.005)
perc. lung cancer incidence (reg.)	-0.002 (0.002)	-0.002 (0.002)	0.002 (0.006)	0.002 (0.006)
perc. diabetes (reg.)	-0.005*** (0.002)	-0.005*** (0.002)	-0.003 (0.008)	-0.003 (0.007)
perc. hypertension (reg.)	-0.003* (0.002)	-0.003* (0.002)	-0.004 (0.008)	-0.004 (0.007)
perc. bronchitis (reg.)	0.006*** (0.002)	0.006*** (0.002)	0.002 (0.006)	0.002 (0.006)
perc. allergies (reg.)	0.004*** (0.001)	0.004*** (0.001)	0.005 (0.005)	0.005 (0.005)
perc. heart cond. (reg.)	-0.002 (0.001)	-0.002 (0.001)	0.003 (0.006)	0.003 (0.006)
perc. smokers (reg.)	-0.006*** (0.001)	-0.006*** (0.001)	-0.006 (0.006)	-0.006 (0.006)
perc. ICU beds 50km radius	-0.004*** (0.001)	-0.004*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)
cases vs tests ratio (reg.)	0.044*** (0.001)	0.044*** (0.001)	0.045*** (0.001)	0.045*** (0.001)
median density	0.003** (0.001)	0.003** (0.001)	0.002** (0.001)	0.002** (0.001)
perc. females	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
perc. over 65	-0.001 (0.001)	-0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
over 74 conc.	0.003*** (0.001)	0.003*** (0.001)	0.002*** (0.001)	0.002*** (0.001)
perc. working out of comune	-0.009*** (0.001)	-0.009*** (0.001)	-0.010*** (0.001)	-0.010*** (0.001)
perc. families over 4	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
perc. employed	-0.001 (0.001)	-0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
perc. degree (over workforce)	-0.002** (0.001)	-0.002** (0.001)	-0.002 (0.001)	-0.002 (0.001)
perc. low income	-0.005*** (0.002)	-0.005*** (0.002)	-0.007*** (0.002)	-0.007*** (0.002)
log. agedcare employees	0.007*** (0.001)	0.007*** (0.001)	0.006*** (0.001)	0.006*** (0.001)
log. construction employees	0.002 (0.003)	0.002 (0.003)	0.003 (0.003)	0.003 (0.003)
log. delivery employees	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)
log. education employees	0.002 (0.003)	0.002 (0.003)	-0.0003 (0.003)	-0.0002 (0.003)
log. healthcare employees	0.010*** (0.002)	0.010*** (0.002)	0.009*** (0.002)	0.009*** (0.002)
log. hospitality employees	-0.024*** (0.003)	-0.024*** (0.003)	-0.023*** (0.003)	-0.023*** (0.003)
log. manufacturing employees (exc. meat)	0.004* (0.002)	0.005* (0.002)	0.005** (0.002)	0.005** (0.002)
log. manufacturing employees (meat only)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
log. other employees	-0.012*** (0.002)	-0.012*** (0.002)	-0.013*** (0.002)	-0.013*** (0.002)
log. retail employees	-0.006 (0.004)	-0.006 (0.004)	-0.006 (0.004)	-0.006 (0.004)
log. services employees	0.010*** (0.004)	0.010*** (0.004)	0.012*** (0.004)	0.012*** (0.004)
log. transportation employees	-0.003 (0.002)	-0.003 (0.002)	-0.002 (0.002)	-0.002 (0.002)
log. warehousing employees	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
log. wholesaling employees	0.028*** (0.003)	0.028*** (0.003)	0.028*** (0.003)	0.027*** (0.003)
log. hotel beds	0.015*** (0.001)	0.015*** (0.001)	0.015*** (0.001)	0.015*** (0.001)
tot. length primary roads	0.016*** (0.001)	0.016*** (0.001)	0.017*** (0.001)	0.017*** (0.001)
past excess mort. X perc. lockdown	-0.025*** (0.002)	-0.025*** (0.002)	-0.025*** (0.002)	-0.025*** (0.002)
past excess mort. X neighb. excess mort.	0.010*** (0.0001)	0.010*** (0.0001)	0.010*** (0.0001)	0.010*** (0.0001)
neighb. excess mort. X perc. lockdown	-0.031*** (0.002)	-0.031*** (0.002)	-0.032*** (0.002)	-0.031*** (0.002)
past excess mort. X neighb. excess mort. X perc. lockdown	-0.001** (0.0003)	-0.001** (0.0003)	-0.001** (0.0003)	-0.001** (0.0003)
Constant	0.013*** (0.001)	0.013*** (0.001)	0.009** (0.004)	0.009** (0.004)
Observations	1,112,521	1,112,521	1,112,521	1,112,521
R ²	0.355	0.355		
Adjusted R ²	0.355	0.355		
Log Likelihood			-1,372,881.000	-1,372,884.000
Akaike Inf. Crit.			2,745,872.000	2,745,878.000
Bayesian Inf. Crit.			2,746,528.000	2,746,534.000
Residual Std. Error (df = 1112468)	0.831	0.831		
F Statistic (df = 52; 1112468)	11,794.850***	11,794.830***		

Note:

*p<0.1; **p<0.05; ***p<0.01

3.2 Standardised coefficient plot



3.2.1 Top 20 coefficients

label	Estimate	Estimate_low	Estimate_upp
past excess mortality	0.4097	0.4083	0.4110
neighbouring excess mortality	0.2087	0.2070	0.2097
cases vs tests ratio (reg.)	0.0446	0.0434	0.0453
log. wholesaling employees	0.0276	0.0243	0.0304
nitrogen dioxide conc.	0.0205	0.0164	0.0215
tot. length primary roads	0.0165	0.0153	0.0174
log. hotel beds	0.0148	0.0137	0.0158
log. services employees	0.0112	0.0064	0.0142
temperature	0.0098	0.0083	0.0109
log. healthcare employees	0.0095	0.0070	0.0111
carbon monoxide conc.	0.0067	0.0041	0.0078
log. agedcare employees	0.0064	0.0048	0.0070
log. manufacturing employees (exc. meat)	0.0049	0.0020	0.0069
PM2.5 conc.	0.0045	0.0027	0.0054
perc. allergies (reg.)	0.0045	0.0005	0.0047
perc. families over 4	0.0040	0.0023	0.0048
log. manufacturing employees (meat only)	0.0037	0.0026	0.0046
log. delivery employees	0.0034	0.0013	0.0055
PM10 conc.	0.0029	0.0018	0.0037
over 74 conc.	0.0027	0.0014	0.0032

3.3 Predictions for the Province of Bergamo

Predictions are based on the pooled model (with PM 2.5) and limited to the cells of the Province of Bergamo.

Predictions details:

- Predictions are produced for each cell of the Province of Bergamo and for each day between 1 February 2020 and 30 April 2020.
- Predictions are set with total lockdown measures implemented since 11 February 2020.
- The missing values for the case-to-test-ratio variable are imputed based on the first regional value available (reported for 24 February 2020, when testing started in Italy)
- To include the effect of *past* and *neighbouring* excess mortality levels are produced in three steps:
 1. Daily predictions for the 33 cells of the Province are run based on observed data, with population under strict lockdown measures set to 100%.
 2. The resulting daily excess mortality predictions for each cell are used as neighbouring excess mortality predictors for a daily new prediction.
 3. This last daily prediciton is used to update to 4-day lagged local excess mortality predictor which is the used in the predictions for the following day.

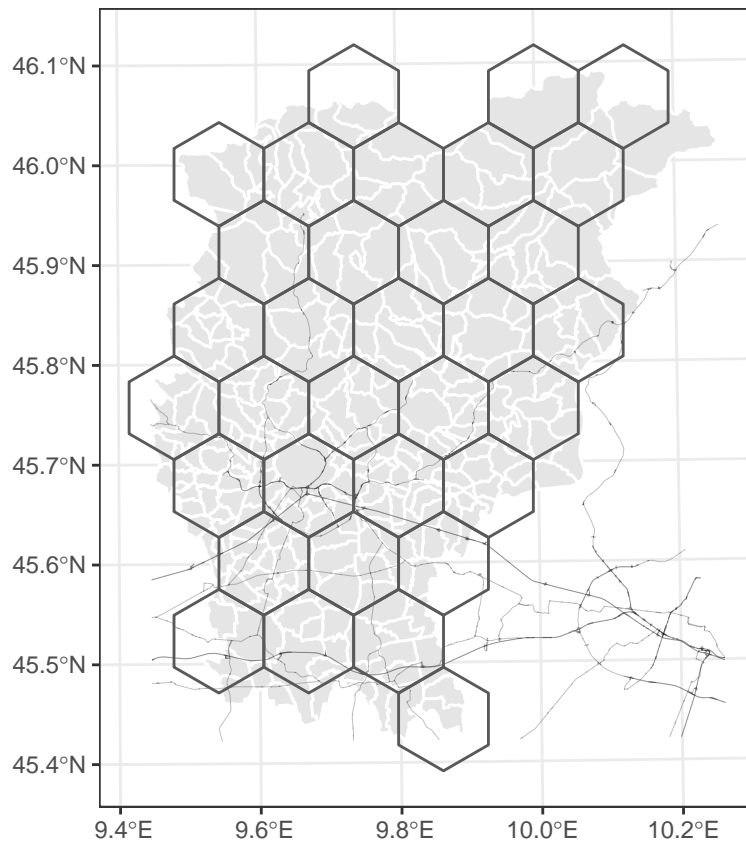


Figure 1: Comuni in the Province of Bergamo and major roads

